**Deep papers**

[Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift](http://arxiv.org/pdf/1502.03167v3.pdf)

For deep network, if only input layer and output layer are whitened, during training, hidden layers are gradually deviated from zero mean, unit variance and decorrelated conditions. We refer to the change in the distributions of internal nodes of a deep network, in the course of training, as *Internal Covariate Shift.* Batch normalization comes to address that issue.